

重力モデルを用いたサッカー選手の動きの定量化

土田 潤¹・宿久 洋²

(受付 2016 年 12 月 30 日；改訂 2017 年 3 月 10 日；採択 5 月 15 日)

要 旨

サッカー選手の動きに関して、ボールを保持している選手に着目した研究は多数存在する。また、ボールを持っていない選手を含めた研究は戦術などの、選手全体の動きに関する評価が多く、個人への評価は少ないのが現状である。本稿では、全体の動きを考慮しつつ、選手個人の動きを評価する指標として、選手質量を用いる。選手質量の定義には重力モデルを用いる。重力モデルにおいて選手質量はパラメータとなるが、重力モデルの式が対数線形モデルと同様のモデル式として定義されることから、選手質量の推定が可能となり、選手質量について、対数線形モデルにおける主効果と同様の意味付けができることを報告する。また、本稿ではトラッキングデータを用いて、選手間密度、および選手間距離を算出し、選手質量を推定した結果を報告する。加えて、推定された選手質量に対して、階層ベイズモデルを用いてどのようなプレーが選手質量に影響を与えているかをモデリングした結果を報告する。

キーワード：階層ベイズモデル、スポーツデータ解析、対数線形モデル。

1. はじめに

サッカーは世界的に非常に人気のスポーツであり、多くのプロチームが存在する。プロチームは勝利のためにデータを活用することが多く、データ活用の論文も多く発表されている。例えば、Rue and Salvesen (2000)では、ゲームのスコアを予測するモデルをベイジアン Dynamic 一般化線形モデルを用いて構築している。Dobson and Goddard (2008)では、ホーム・アウェイなどの情報を用いることで、ゲームスコアの推定を行っている。これらの研究に代表されるように、多くの研究は各試合の集計結果を用いたデータを活用することが主流である。また、サッカー選手の動きやサッカーのシーンをどのように観測するかも研究対象となっている。Ekin et al. (2003)や Ren and Jose (2005)では、サッカー中継から、サッカー中継のどのようなシーンなのかを判別するモデルを提案しており、選手の動きを画像検出の方法を用いて類型化し、ズーム、リプレイなどのシーンを判別している。

サッカーのチーム戦術の分野では、ロボットサッカーにおいて研究が盛んである。その理由としてロボットサッカーは、4人チームであり、選手がロボットであることから個々の能力の差が少ないため、人対人に比べ、戦術によって勝敗が決定しやすいことが挙げられる。また、フィールドの大きさが人対人のサッカーに比べ小さく、カメラ一台でフィールド全体を写すことができ、データの観測が比較的容易であることも戦術研究が盛んになる要因であると考えられる。Nadarajah and Sundaraj (2013)では、ロボットとボール間の距離やそれぞれの速度を用

¹ 同志社大学大学院 文化情報学研究科：〒610-0394 京都府京田辺市多々羅都谷 1-3

² 同志社大学 文化情報学部：〒610-0394 京都府京田辺市多々羅都谷 1-3

いて、ロボットのポジショニングに関する戦術や、ゴールをどのように守ればよいのかを総合的に報告している。

しかし、近年の観測技術の発達により、人対人のサッカーにおいても、ボールを持った選手だけでなく、ボールを持っていない選手に関するデータが得られるようになり、ある時点でのフィールドにいる選手全員のプレーを分析することが可能になった。全員を同時に、特に、ボールを持っていない選手を分析することは重要である。加藤(1999)では、サッカーにおいて1人の選手がボールをプレーするのは多くても2分から3分であり、選手がボールを持っていない状態での動きの解析が求められていることが述べられている。

実際に、ボールを持っていない選手を含めたサッカーの分析は近年多く行われている。神谷他(2016)では、選手のトラッキングデータを用いて、サッカーの戦況変化の自動検出を選手の守備位置や選手間の距離を用いて行なわれている。また、成塚 他(2016)では、選手間の距離や角度の特徴について考察が行われている。可視化についてもボールの動きだけでなく、船山他(2016)では、ホットゾーンの利用人数の動的表示やカーネル密度推定を用いてフィールドのどこに人が集中しているかを可視化している。これらの研究の多くは、フィールド全体ないし、選手全体について考察していることが多く、個人の能力について言及しているものは少ない。

個人の能力や戦術の評価については、トラッキングデータではなく、ビデオデータを用いて行われている研究がある。瀧・長谷川(1998)では、チームという集団の動きの中で個人が作る領域について考察し、ボロノイ領域を拡張することで個人の特徴量を算出し、得られた特徴量から個人の動きを評価している。特に、サッカーという限られたエリアでの集団の密度は、各個人の周りの密度の集合として還元されると考えられることを述べている。つまり、選手個人の動きによって作られる密度を評価をすることで、全体の密度が評価できるとしている。また、Moura et al. (2015)では2012年ヨーロッパチャンピオンシップのデータに対して、個人の動きに対してPCAを適用し、選手のポジションやその範囲がチームごとによって異なることを示している。このように、選手全員を用いた、選手個人の動きに対する定量的評価はまだ初期段階であると考えられる。

本稿では、瀧・長谷川(1998)の考え方にに基づき、個人の周りの密度を評価することで、全体を考慮しながら、選手個人の動きを評価することを考える。まず、選手間の範囲で作られる密度はその選手個人の質量によって表現されるものとして捉え、選手個人の質量を選手間で構成される範囲の密度を表すパラメータとして重力モデルを用いて定式化する。このとき、選手間の範囲で作られる密度はデータとして与えられるが、個人の質量はパラメータである。そのため、データから個人質量を推定する必要があるが、個人質量の推定が、対数線形モデルの主効果の推定と同様であることを報告する。

本稿では、実際のトラッキングデータを用いて、個人質量を推定し、さらに、推定された個人の質量を高める行動とはどのような行動であるのかを階層ベイズモデルによって表現した結果を報告する。2章では、重力モデルを用いるためのデータの作成法および、個人質量の推定法について述べる。3章では、推定された個人の質量について階層ベイズモデルを適用し、どのような行動が質量に関連しているかを明らかにする。4章では、まとめとして、得られた結果に対する考察および、今後の課題について述べる。

2. 重力モデルを用いた選手の質量推定

この章では重力モデルを用いて選手の質量を定式化する。そして、選手質量の推定が、対数線形モデルの推定法と同一となることから、選手質量の意味について考察する。

2.1 重力モデルを用いた選手質量の定式化

重力モデルは Tinbergen (1962) に代表されるように経済貿易のモデルとして用いられている。簡略化された重力モデルは対象 i, j 間で観測される量を F_{ij} とすれば,

$$(2.1) \quad F_{ij} = \frac{M_i M_j}{D_{ij}}$$

で定義される。ここで、 D_{ij} は対象 i, j 間の距離であり、 M_i は対象 i の質量である。両辺の対数を取れば,

$$(2.2) \quad \log F_{ij} = \log M_i + \log M_j - \log D_{ij}$$

と表される。経済貿易のモデルでは、 F_{ij} を国 i, j 間の貿易量、 D_{ij} を国 i, j 間の距離、 M_i を国 i の総貿易量としている。

本稿では F_{ij} を選手 i, j 間の密度とする。 F_{ij}, D_{ij} が与えられたもとで、 $\log F_{ij}^* = \log F_{ij} + \log D_{ij}$ と置けば、式 (2.2) は

$$(2.3) \quad \log F_{ij}^* = \log M_i + \log M_j$$

と与えられる。サッカーのフィールドの大きさは状況によらず一定であり、多くの場合、フィールドに計 22 人の選手でいることから、質量の総和は一定であるとする。

重力モデルの良い点として、選手間密度および、選手間距離が与えられれば、選手質量を定義することができる点が挙げられる。また、 i, j を選手だけでなく、フィールドを分割した区分とすれば、フィールドの質量として解釈も可能である。質量を解釈すれば、質量が大きい選手や区分の周辺に選手が集まっていることを意味する。

本稿では、図 1 のような選手間領域と選手領域を用いて、選手間密度 F_{ij} を定義する。選手 i, j 間の距離 D_{ij} を半径、対象 i を中心とする円を選手 i, j 間の領域 R_{ij} とし、 S_k を i の対戦チームの選手 k の領域とする。 S_k は i の対戦チームの選手 k を中心とする半径 1.7m の円とした。領域の半径 1.7m としたのは成人の平均身長であり、選手が一步、領域方向に進み、手を伸ばせば領域 R_{ij} に侵入できるためである。このとき、 R_{ij} と S_k が交わる k の個数を R_{ij} の面積で割ったものを F_{ij} とする。つまり、 F_{ij} は以下の式で与えられる。

$$(2.4) \quad F_{ij} = \frac{\sum_{k \in O_i} I(R_{ij} \cap S_k \neq \phi)}{D_{ij}^2 \pi}$$

ここで、

$$R_{ij} = \{x \in \mathbb{R}^2 \mid \|x - y_i\| \leq D_{ij}\}$$

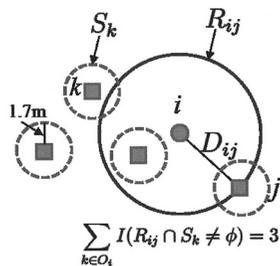


図 1. カウントする領域と選手の持つ領域。

$$S_k = \{x \in \mathbb{R}^2 \mid \|x - y_k\| \leq 1.7\}$$

であり、 O_i は対象 i の対戦チームの選手集合、 y_i は対象 i の座標ベクトルである。また、 $I(\cdot)$ は定義関数であり、 π は円周率である。式(2.4)を F_{ij}^* で書き直せば、

$$(2.5) \quad F_{ij}^* = \frac{\sum_{k \in O_i} I(R_{ij} \cap S_k \neq \phi)}{D_{ij} \pi}$$

となる。経済で用いられている重力モデルでは F_{ij} は対称であることが多いが、式(2.4)および、式(2.5)は非対称、つまり、 $F_{ij} \neq F_{ji}$, $F_{ij}^* \neq F_{ji}^*$ となることに注意されたい。

これは、中心が異なる円を用いている点、および、対戦相手の数を数えているためであり、領域 R_{ij} で数え上げる対象選手をフィールドにいる選手全員としても、必ずしも対称となるわけではない。また、対象 i の座標ベクトルを2次元としたのは、選手座標が X, Y 座標によって取られていることを想定しているためであり、一般に2次元である必要はない。

式(2.5)を用いて選手間密度を定義した場合、選手質量が大きい選手には、近くに多くの選手が存在しやすいことを表す。つまり、分母である選手間距離が小さく、分子である近くにいる選手の数が多い傾向があると考えられる。攻撃側の視点に立てば、質量の大きい選手は相手に厳しくマークされていることが読み取れ、質量が小さい選手はフリーであることが読み取れる。相手のゴールに近い位置で質量が大きい選手は囷として、フリーになる選手を作る役割を持つことができる。また、相手のゴールに近い位置で質量が小さい選手は、フリーに近い状態であると考えられるため、その選手がボールを得ることができれば、得点を期待することができる。守備側の視点に立てば、質量の大きい選手は近くに相手選手がおり、相手がフリーになるのを防いでいると考えることができる。一方で、質量が小さい選手は近くに相手選手が少なく、その選手自体がフリーの状態であり、フリーという意味で、守備に積極的に貢献しているわけではないと考えられる。

2.2 選手質量の推定

式(2.3)は和について制約を入れれば、対数線形モデルと同様となる。ここでは、対数線形モデルと同様、 $\log F_{ij}^*$ を i, j についての総和を0とし、それぞれの選手質量の総和を0とする。 $\log F_{ij}^*$ の総和が一定というの、フィールドの大きさおよび、選手の数が一定であることから、自然な仮定である。以上のことからモデル式は以下ようになる。

$$(2.6) \quad \log F_{ij}^* = \log \lambda_i + \log \kappa_j + \mu$$

ただし、

$$\sum_{i \in O} \sum_{j \in O} \log F_{ij}^* = 0, \quad \sum_{i \in O} \log \lambda_i = 0, \quad \sum_{j \in O} \log \kappa_j = 0$$

である。ここで、 O はフィールドにいる選手の集合である。パラメータの推定量は対数線形モデルと同様、

$$\mu = \frac{1}{|O|^2} \sum_{i \in O} \sum_{j \in O} F_{ij}^*, \quad \log \lambda_i = \frac{1}{|O|} \sum_{j \in O} F_{ij}^* - \mu, \quad \log \kappa_j = \frac{1}{|O|} \sum_{i \in O} F_{ij}^* - \mu$$

で与えられる。 μ は総和を一定にするパラメータであり、どれだけフィールドにフリーである領域が残っているかを表すパラメータと解釈することができる。 $\log F_{ij}^*$ を i, j についての総和を0というデータの制約のため、 μ は0として推定される。実際には、与えられたデータである $\log F_{ij}^*$ の平均を求め、 $\log F_{ij}^* - \mu$ を新しいデータとし、その行平均、列平均を用いて選手質

量を求めればよい。

式(2.6)を対数線形モデルとして捉えれば、選手質量の対数は対数線形モデルの行主効果、列主効果としてとらえることができる。また、対数線形モデルであることから、対象 i に対して行主効果、列主効果の2つが推定される。対象 i の行主効果および列主効果の値は F_{ij}^* が非対称であることから、一般には異なる。つまり、 $\log \lambda_i \neq \log \kappa_i$ である。そのため、2種類の選手質量の対数 $\log \lambda_i, \log \kappa_i$ はそれぞれ解釈が異なる。行主効果の推定値 $\log \hat{\lambda}_i$ は選手 i を中心とする円の平均的な密度を表している。したがって、行の質量が大きい選手は周りに相手チームの選手が多く、その選手の周りにスペースが少ないことを表している。したがって、選手 i がどれだけ相手チームの選手にマークされているかがわかる。選手質量として解釈すれば、行の質量はその選手自身が相手選手をどれだけ引き付けているかを示す量であり、大きい選手の周りに他の選手が集まりやすいことを示す。

列主効果の推定値 $\log \hat{\kappa}_i$ は相手チームの選手を中心とし、選手 i との距離を半径とする円の平均的な密度について表している。そのため、 $\log \hat{\kappa}_i$ が大きい選手は、相手チームを中心とした円の半径が小さくなりやすい位置にいる傾向がある。このことから、列の質量が大きい選手は相手チームの選手との重心にいることを示していると考えられる。つまり、守備の意味では、相手チームの重心にいることによりボール回しを止められる位置にいると考えられ、攻撃の意味では、動くことで、どちらにスペースを作るかを決められるポジションにいると考えられる。選手質量として解釈すれば、質量が大きい選手は行側の選手質量の釣り合いを保つための質量が大きいことを意味している。

3. 階層ベイズモデルを用いた選手質量と選手の動きのモデル

本章では、2016年スポーツデータ解析コンペティションで提供されたデータスタジアム提供のトラッキングデータ及びボールタッチデータから、2章で定義した重力モデルを用いて選手質量を推定する。そして、ボールタッチと推定された選手質量との関連を階層ベイズモデルによってモデリングした結果を報告する。

3.1 データ概要およびデータの整形

分析対象として、2016年スポーツデータ解析コンペティションで提供されたデータスタジアム提供のトラッキングデータ及びボールタッチデータを用いる。それぞれ、2016年2月27日に行われた磐田対名古屋、広島対川崎、鳥栖対福岡、柏対浦和、湘南対新潟、神戸対甲府、横浜FM対仙台、FC東京対大宮、および2016年2月28日に行われたG大阪対鹿島の計9試合のデータであり、トラッキングデータでは、1/25秒ごとに各選手の X, Y 座標を計測している。ボールタッチデータでは、ボールタッチが起こった際の、ボールタッチを行った選手の X, Y 座標および、アクションが1/30秒のフレーム単位によって記述されている。ボールタッチデータは、あくまで、ボールタッチが起きたときのみ記録される。

ボールタッチデータとトラッキングデータは観測されているフレームレートが異なるため、ボールタッチデータを一度秒に直し、再度1/25秒のフレームに直した。端数は四捨五入し、トラッキングデータとフレーム数が同じ場合のみ結合を行った。また、 X, Y 座標はトラッキングデータに合わせ、ボールタッチが行われていない場合のボールの位置 X_t, Y_t を以下のように線形補完した。つまり、

$$X_{t+1} = X_t + a_x, Y_{t+1} = Y_t + a_y \quad (t = 0, 1, \dots, T-1)$$

とした。ここで、 a_x, a_y は変化量であり、

表 1. 階層ベイズモデルに用いた説明変数.

変数名	説明
選手位置	選手位置に関するダミー変数 (図 2)
ボール位置	ボール位置に関するダミー変数 (図 2)
累積パス回数	その時点までに行ったパスの総回数
累積シュート回数	その時点までに行ったシュートの総回数
ボールとの距離	ボールとのユークリッド距離. 単位は図 2 での座標の単位
時速	選手の移動速度

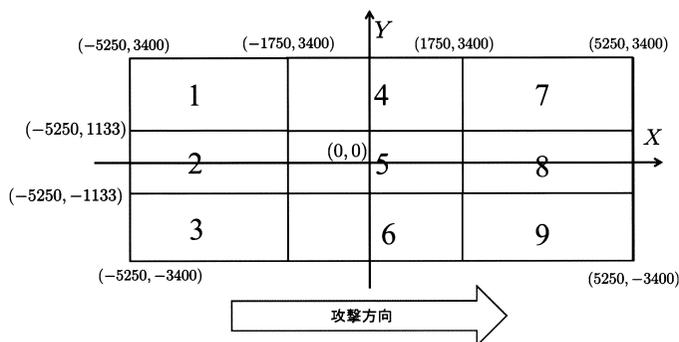


図 2. 計測したフィールドのスケール及びボール分割の位置について.

$$a_x = (X_0 + X_T)/T, \quad a_y = (Y_0 + Y_T)/T$$

である. X_0, Y_0 は観測されているボールの X, Y 座標であり, X_T, Y_T は X_0, Y_0 の次に観測されているボールの X, Y 座標である. T は X_0, Y_0 と X_T, Y_T が観測される間の総フレーム数である.

どのような行動が選手質量と関わるかを調べるため, 説明変数として, 行動に関わる変数を選択した. 用いた変数は表 1 のとおりである. 選手位置, ボール位置は図 2 のようにそれぞれフィールドを 9 分割したものに, その位置に選手, またはボールがいた場合 1, そうでない場合は 0 とするダミー変数を用いた.

これらの説明変数は各フレームで観測されている. 例えば, ボールとの距離では, 各時点で, 選手の位置と線形補完されたボールの位置で算出されている. また, ボール位置および選手位置は, 攻撃方向によって自陣か敵陣かが異なるため, 攻撃方向をそろえ, 算出している.

目的変数は, 各時点での目的変数としては, 対数線形モデルの行主効果, つまり選手質量の対数を用いる. しかし, i, j 間に相手チームの選手がいない場合, $\log F_{ij}^*$ の値が $-\infty$ となることから, 今回は, F_{ij}^* の数え上げを行う際に, チームごとにデータを分割した. つまり, 各チームごとに自チームの選手を行, 相手チームの選手を列にした密度行列を各フレームごとに作成し, 選手質量の推定を行った. 対象となる選手はゴールキーパーを含むフィールドにいる選手全員である. したがって, 退場などの特別な場合を除き, 密度行列の大きさは 11×11 である.

今回の分析では, 行主効果について着目した. 目的変数の値の例として, 図 3 に示す. 縦軸は, 推定された行主効果, つまり, 選手質量の対数を表している. 行主効果として解釈すれば, 0 以上では相対的に質量が大きく, 相手選手を引き付けているということを表す. また, 横軸は, フレーム数であり, 25 フレームで 1 秒を表す. 図 3(a) は川崎の FW である大久保選手の推定された行主効果, つまり選手質量の対数であり, 図 3(b) は広島 MF の茶島選手の選手

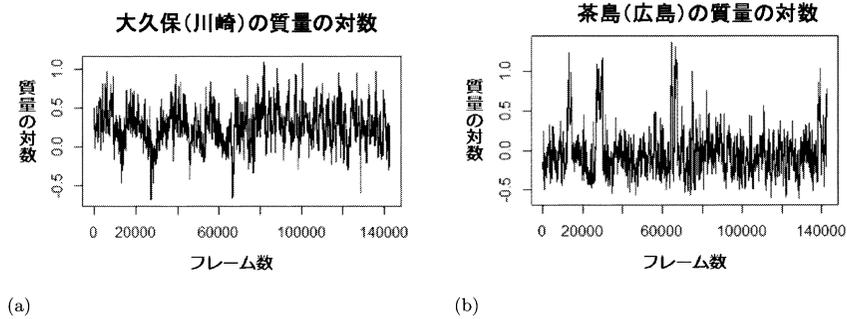


図 3. 推定された選手質量の対数の例.

質量の対数である。大久保は、全体的に 0 以上の値を取っており、前線の選手として、相手の DF にマークされていることがわかる。大久保は前年の得点王であり、また、この試合が、第 1 節であることからマークが厳しいことは、直観的である。また、時折、0 未満の値を取っているが、このとき、川崎のゴール前での攻防が行われていたことから、相手チームの選手が大久保を無視して、攻撃参加していたと考えられる。

一方、茶島は全体的に 0 未満の値を取っている。MF というポジションであることから、相手選手に張り付くことより、どの相手にボールがわたってもすぐに行動できるようなポジションを取っているため、行主効果が低く推定されたと考えられる。また、一時的に大きな値を取っているが、このときは相手ゴール前での攻防であり、ゴール前へのパスに反応するために茶島がゴールまで待機している場面であった。

3.2 階層ベイズモデル

前節で見た通り、選手質量の値が大きくなる場面と小さくなる場面がポジションによって異なると考えられる。FW のような攻撃的なポジションでは攻撃の際に質量は大きくなりやすく、一方で、DF や GK のような守備的なポジションでは守備の場面において質量が大きくなりやすいと考えられる。同様に、チームによっても異なると考えられる。これは、各チームにおいて、戦術が異なることや、チームが選手にどのような役割を与えているかによって、質量を大きくすることが目的となる選手も存在すれば、質量を小さくすることが目的の選手も存在するからである。さらに、前半後半によっても行動の質が異なると考えられる。特に後半は交代が行われやすく、選手は前半と異なる行動をとると考えられる。以上より、チーム、選手、前後半によって回帰係数の値が異なると考え、階層モデルを適用する。また、途中出場の選手は相対的にサンプル数が少ないため、ベイズモデルを用いる。

適用したモデルは Rossi et al. (2005) のモデルであり、以下で表現される。

$$\begin{aligned}
 (3.1) \quad & \begin{cases} y_{it} = \mathbf{x}'_{it}\beta_i + \varepsilon_{it} \\ \beta_i = \Theta \mathbf{z}_i + \mathbf{v}_i \end{cases} \\
 & \varepsilon_{it} \sim N(0, \tau_i) \\
 & \tau_i \sim n_e s q_i / \chi^2(n_e) \\
 & \mathbf{v}_i \sim N(0, \mathbf{V}_\beta) \\
 & \text{vec}(\Theta) | \mathbf{V}_\beta \sim N(\text{vec}(\hat{\Theta}), \mathbf{V}_\beta \otimes \mathbf{A}^{-1})
 \end{aligned}$$

表 2. 階層の説明変数.

説明変数名	説明
チームフラグ	各チーム ID を表すダミー変数
前後半フラグ	前半後半を表すダミー変数

$$V_{\beta} \sim IW(n_{\beta}, V)$$

ここで, $n_e, n_{\beta}, sq_i, V, A, \Theta$ はそれぞれハイパーパラメータである. ハイパーパラメータの設定には, Rossi et al. (2005) に則し, $n_e = 3, n_{\beta} = k + 3, sq_i = s_i^2, V = n_{\beta}I, A = 0.01I, \text{vec}(\Theta) = 0$ とした. ここで, $y_{it} \in \mathbb{R}$ ($i = 1, 2, \dots, n, t = 1, 2, \dots, T_i$) は目的変数であり, $x_{it} \in \mathbb{R}^p$ は説明変数, $z_i \in \mathbb{R}^k$ は個人の属性を表す階層の説明変数であり, $\beta_i \in \mathbb{R}^p$ は対象 i に対する回帰係数ベクトル, $\Theta \in \mathbb{R}^{p \times k}$ の階層に対する回帰係数行列である. n は対象数であり, T_i は対象 i の総フレーム数である. s_i^2 は対象 i の目的変数の標本分散であり, I は単位行列である. また, $\text{vec}(\cdot)$ は vec オペレータ, \otimes はクロネッカー積を表す.

説明変数の一覧は表 1 である. 目的変数は各時点での対象 i の推定された行主効果を用いた. 階層に用いた変数は表 2 の通りである. したがって, 対象 i はそれぞれの選手と前半後半の組み合わせとなる. ただし, 累積パス回数および累積シュート回数はその試合全体で算出した. また, 各対象でキックオフの時点から 5 秒まではデータから除外した. また, 総フレーム数が 1250 フレーム以下の対象も除外したデータ行列を作成した. このとき, 対象数は 444 となった.

事後分布の推定には, Gibbs sampler を用いた. 総サンプル数は 200000 回とし, 焼き入れ数を 100000 回とした. また, 初期値はランダムに発生したものを 3 回行った. 収束判定には目視による確認と, Gelman and Rubin (1992) の方法を用いて, \hat{R} 値が 1.1 未満であるかを基準として確認した.

β_i の推定値には事後平均値を用いた. 各 i の事後平均値のヒストグラムを図 4 に示す. 図 4 の縦軸は度数であり, 横軸は事後平均値である. 選手位置_1 および, ボール位置の 1 がないのは, ダミー変数の基準として用いたためである.

図 4 から, 選手位置_8 の係数が大きいものが多いことがわかる. つまり, 相手ゴール前の付近にいる選手の質量は大きくなりやすいことがわかる. また, 自陣ゴール前の付近である選手位置_2 も同様に大きくなりやすい傾向がみられる. これは, DF や GK などのゴールに関わる位置にいる選手は相手選手を引き付ける, もしくは引っ張られやすいということが読み取れる. また, 累計パス回数の回帰係数のヒストグラムを見れば, 回数を増やせば質量が大きくなる選手とそうでない選手, つまり, 回帰係数が正である選手と負である選手が半分に分かれている. このような結果となったのは, ポジションによる影響ではないかと推察される. ボールとの距離の回帰係数の値が小さいが, これは, 単位が他の変数と異なり, 説明変数に用いた値が大きいためである. また, 回帰係数の値が負である選手が多いことから, ボールとの距離の値が大きければ大きいほど, 選手質量が小さくなる傾向があることが読み取れる. これは, ボールを持っている選手をフリーにすることは危険であるという認識が推定結果として現れたのではないかと考えられる. 時速の推定値のヒストグラムを見れば, 素早い動きをする選手の質量が大きくなりやすい傾向が見て取れる. このことから, 速い動きにつられ, 相手選手が移動していると考えられるが, そもそも, 素早い動きをしなければならぬ場面が, 質量が大きくなりやすい場面であるとも考えられる.

以上から, 相手選手を引き付ける方法として, 相手ゴール正面のエリアで素早い動きを行えば, 一部 DF はつられ, 囷として機能するのではないかと解釈できる. フィールドが一定である

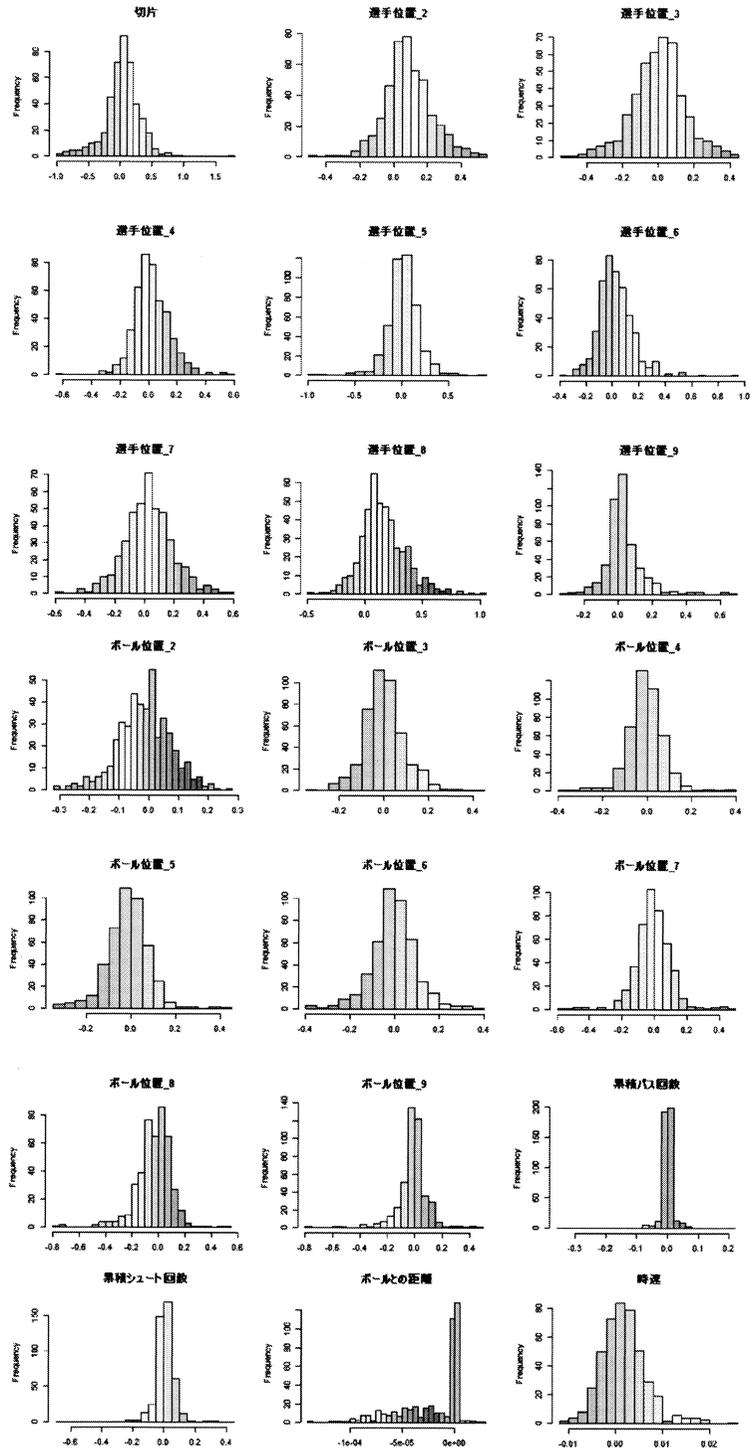


図 4. 回帰係数の事後平均値のヒストグラム.

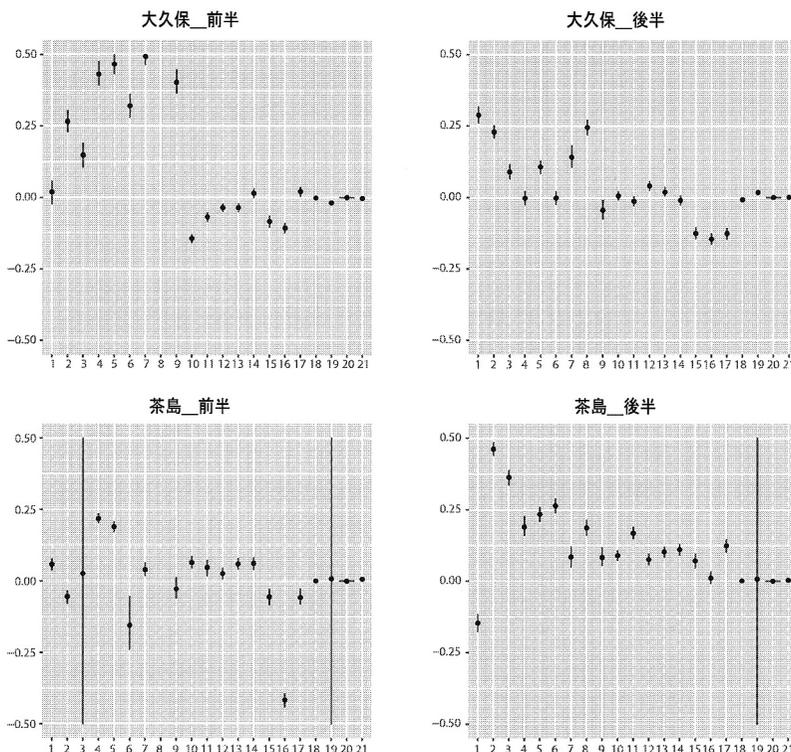


図 5. 大久保および茶島の回帰係数の事後分布。

ことから、一部の選手の付近に他の選手が集まれば、その他の選手がフリーになりやすい。そのため、ゴール付近に、一人は素早い動きで相手を引き付け、もう一人は相対的にゆっくりと動けば、相対的にフリーを作れるのではないかと考えられる。

次に、個人の回帰係数の事後分布について考察を行う。図 5 のバイオリンプロットは大久保および、茶島の回帰係数の事後分布である。縦軸は回帰係数の値であり、横軸の番号と説明変数は表 3 のように対応している。バイオリンプロットの黒点は事後平均値の値である。表 4 はそれぞれの事後平均を表している。

大久保、茶島それぞれ、前半と後半で異なる傾向を表している。大久保の前半では、どの選手位置にいても質量が上昇する傾向にあるが、後半になれば、一部の位置で 0 に近い値、または負の値を取るようになっていく。しかしながら、切片の値が大きく、依然として、相手を引き付ける力が高いことがわかる。累積シュート回数の係数に着目すれば、前半では、シュート回数が増えても、質量が増加する傾向はみられないが、後半はシュート回数が増えれば、質量が上昇する傾向があることがわかる。

茶島は、前半は選手位置によって質量が上昇する傾向や、低い傾向が散見されるが、後半は、どの位置にいても、どの位置にボールがあったとしても、ボール位置_1 に比べ、選手質量が上昇する傾向がある。しかしながら、後半の切片が負の値を取るから、平均的な質量は大きくなりにくいことがわかる。

茶島の累計シュート回数および、選手位置_2 の値はほとんどが 0 であるため茶島の累計シュート回数および、選手位置_2 の回帰係数の事後分布の分散が大きい。

表 3. 変数番号と説明変数の対応表.

番号	説明
1	切片
2~9	選手位置 2~9
10~17	ボール位置 2~9
18	累積パス回数
19	累積シュート回数
20	ボールとの距離
21	時速

表 4. 大久保および茶島の回帰係数の事後平均.

	大久保 前半	大久保 後半	茶島 前半	茶島 後半
切片	0.019	0.288	0.058	-0.146
選手位置_2	0.266	0.229	-0.055	0.462
選手位置_3	0.148	0.090	0.036	0.364
選手位置_4	0.431	-0.002	0.219	0.190
選手位置_5	0.466	0.107	0.189	0.235
選手位置_6	0.321	-0.002	-0.155	0.264
選手位置_7	0.503	0.141	0.039	0.086
選手位置_8	0.620	0.245	0.527	0.187
選手位置_9	0.403	-0.044	-0.028	0.083
ボール位置_2	-0.144	0.005	0.065	0.090
ボール位置_3	-0.069	-0.013	0.048	0.168
ボール位置_4	-0.036	0.041	0.026	0.076
ボール位置_5	-0.035	0.019	0.059	0.103
ボール位置_6	0.015	-0.010	0.061	0.111
ボール位置_7	-0.084	-0.126	-0.056	0.071
ボール位置_8	-0.107	-0.145	-0.417	0.011
ボール位置_9	0.021	-0.126	-0.058	0.124
累積パス回数	-0.002	-0.007	0.001	0.002
累積シュート回数	-0.019	0.018	0.014	0.017
ボールとの距離	0.000	0.000	0.000	0.000
時速	-0.003	0.001	0.006	0.003

次に、階層の説明変数 Z の回帰係数行列である Θ の事後分布について確認する。図 6 は各説明変数に対する回帰係数 Θ の事後分布を表している。縦軸は回帰係数の値であり、横軸の番号と説明変数は表 3 のように対応している。バイオリンプロットの黒点は事後平均値を表している。また、図 6 に、鹿島の係数がないのは、鹿島を基準にしているためである。表 5 は Θ の事後平均を表している。図 6 と同様の理由で鹿島は記載されていない。図 6 の前後半フラグを確認すれば、切片 0 より大きい点に事後平均があり、0 の部分が分布の裾であることから、前半の方が平均的な密度は大きいことが示唆される。これは、前半は後半に比べ選手が比較的元気であり、厳しいマークを付けることが可能であるためと考える。一方で、チームごとでは、切片を除く変数に大きな違いがみられない。そのため、ボールの位置や選手の位置、時速やパスなどの行動によって、選手質量の差がチームごとでは変わらないと考えられる。この結果は、位置や行動による質量の変化はチーム要因でなく、選手個人の影響が大きいためであると考えられる。他の変数に比べ、切片についてはチームごとで異なる傾向がある。これは、

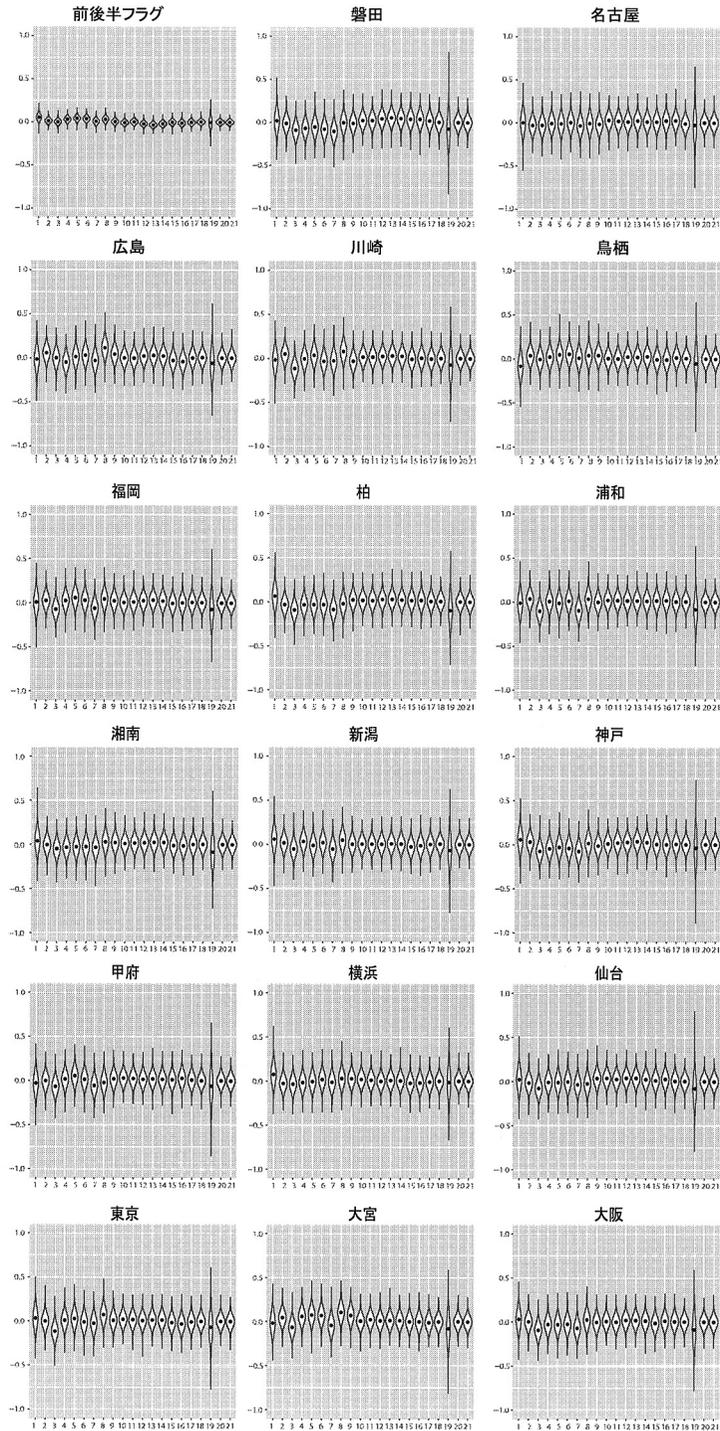


図 6. 階層の説明変数の回帰係数の事後分布.

表 5. 階層の回帰係数の事後平均値.

	前後半フラグ	磐田	名古屋	広島	川崎	鳥栖	福岡	柏	浦和	湘南	新潟	神戸	甲府	横浜	仙台	東京	大宮	大阪
切片	0.053	0.019	0.003	-0.015	-0.017	-0.083	0.007	0.065	-0.011	0.040	0.060	0.057	-0.029	0.076	0.022	0.033	-0.018	0.031
選手位置_2	0.012	-0.011	-0.029	0.059	0.051	0.037	0.027	-0.031	0.037	-0.005	0.018	0.030	-0.001	-0.026	-0.017	0.000	0.045	0.003
選手位置_3	0.001	-0.090	-0.028	0.001	-0.114	-0.006	-0.074	-0.093	-0.102	-0.049	-0.052	-0.075	-0.067	-0.034	-0.079	-0.116	-0.066	-0.093
選手位置_4	0.029	-0.068	-0.007	-0.048	-0.003	0.021	0.025	-0.033	0.007	-0.034	0.033	-0.046	0.018	-0.017	-0.009	0.009	0.058	-0.028
選手位置_5	0.041	-0.052	-0.011	0.015	0.037	0.048	0.058	-0.031	-0.012	-0.028	-0.011	-0.030	0.055	-0.005	-0.011	0.029	0.075	-0.031
選手位置_6	0.036	-0.078	0.004	0.033	-0.032	0.054	0.029	-0.032	0.013	-0.025	0.020	-0.041	0.014	-0.004	-0.004	-0.010	0.069	-0.026
選手位置_7	0.008	-0.103	-0.033	-0.029	-0.026	0.009	-0.061	-0.087	-0.095	-0.034	-0.053	-0.075	-0.054	-0.013	-0.036	-0.024	-0.041	-0.071
選手位置_8	0.031	-0.002	-0.006	0.116	0.081	0.039	0.044	-0.022	0.035	0.031	0.049	0.015	-0.021	0.030	-0.025	0.075	0.106	0.024
選手位置_9	0.004	-0.009	-0.017	0.044	-0.030	0.039	0.023	0.022	-0.001	0.022	0.003	-0.015	0.021	0.028	0.036	0.010	0.071	-0.003
ボール位置_2	-0.008	0.022	0.030	0.001	0.016	0.005	0.004	0.016	0.021	0.010	0.004	0.011	0.030	0.020	0.037	0.022	0.007	0.007
ボール位置_3	0.002	0.023	0.020	-0.001	0.017	0.001	0.010	0.015	0.019	0.016	0.004	0.020	0.025	0.011	0.025	0.017	0.023	0.000
ボール位置_4	-0.021	0.043	0.013	0.023	0.026	0.023	0.024	0.026	0.016	0.020	0.003	0.027	0.016	-0.002	0.039	0.007	0.009	0.020
ボール位置_5	-0.033	0.056	0.025	0.030	0.030	0.019	0.031	0.028	0.015	0.025	0.008	0.037	0.021	0.008	0.037	0.014	0.014	0.019
ボール位置_6	-0.020	0.047	0.012	0.023	0.026	0.027	0.022	0.021	0.012	0.024	0.006	0.026	0.016	0.003	0.023	0.014	0.011	0.012
ボール位置_7	-0.002	0.038	0.010	-0.026	-0.007	-0.008	-0.004	0.016	0.017	-0.011	-0.025	0.004	0.013	-0.023	0.011	-0.015	-0.001	-0.013
ボール位置_8	-0.011	0.039	0.024	-0.040	0.004	-0.009	0.001	0.017	0.018	-0.014	-0.013	0.001	0.029	-0.019	0.027	-0.029	-0.001	0.011
ボール位置_9	-0.003	0.020	0.026	0.003	-0.005	0.013	0.007	0.006	0.004	0.002	0.001	0.006	0.010	-0.008	0.004	-0.007	-0.011	0.004
累積パス回数	0.002	0.005	-0.010	0.007	0.003	0.005	0.005	0.006	0.006	0.002	0.006	0.001	0.001	0.000	0.002	0.000	-0.001	0.003
累積シュート回数	-0.005	-0.073	-0.026	-0.055	-0.072	-0.053	-0.072	-0.097	-0.083	-0.085	-0.065	-0.039	-0.059	-0.013	-0.081	-0.066	-0.078	-0.085
ボールとの距離	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.001
時速	-0.002	0.000	-0.001	0.000	-0.001	-0.003	-0.002	-0.002	-0.001	-0.001	-0.001	-0.001	-0.003	-0.001	-0.003	-0.001	-0.003	-0.001

チームの戦術により、積極的な攻撃をしているかどうかや、前線にいる選手の数に違いがあり、平均的なチームの質量が変改してゐるのではないかと考えられる。

4. おわりに

本稿では、重力モデルを用いて、選手質量を定義し、選手の動きを定量的に評価する方法を提案した。また、定義した選手質量について、トラッキングデータを用いて、どのような行動と関連があるかを階層ベイズモデルによってモデリングを行った。

モデリングの結果から、選手の位置、選手の速度及び、ボールとの距離をうまく利用すれば、フリーの選手を作ることが可能となること等が示唆された。しかしながら、モデリングの際には行側のみに着目し、列側を考慮していない。このことは、選手付近のフリーについてのみモデリングしており、ゴール近く of 攻防を見るのには適しているが、中盤の攻防において、特にパスを回す場面においては適切でない場合がある。

一方で、列側の相手の重心にどうかは、フィールドの中盤の攻防における役割をモデリングする際に必要であると考えられる。また、重力モデルを利用するための密度として今回は単純に相手との距離を利用した円を利用したが、実際には、選手の向きや相手選手の向き、ゴールへの方向など、選手の動きやすい方向によって重みが異なると考えられる。移動方向や向きがわかるのであれば、その向きや速度を用いて重みづけを行い、円形ではなく楕円形によって密度を測ることが考えられる。また、本分析ではゴールキーパーもフィールドプレイヤーとして扱ったが、その性質上、分析対象から除外することも考えられる。さらには、一部のシーンを切り取り、ボールに関わるであろう一部プレイヤーのみを対象とする方法も考えられる。

今回用いたデータでは、1フレームが1/25秒とデータ観測期間が短いため、各フレーム単位で密度行列を作成した今回の分析では、選手質量の自己相関が非常に高い。そのため、推定された選手質量の対数を直接用いるのではなく、一階差分を取った選手質量の対数を用いる方法なども考えられる。また、フレーム単位で密度行列を作るのではなく、シーンごとの場面に区切ることや、秒単位や分単位の代表値を用いる方法なども考えられる。加えて、本分析では、ボールの位置や、ボールとの距離は選手の動きに関連しないことを暗に仮定していたが、実際には、ボールの位置はプレイした選手や近くにいる選手に依存すると考えられる。そのため、ボール位置および、ボール距離をフィールド上の選手位置によって説明するモデルを階層ベイズに組み込むことも考えられる。つまり、ボール位置のダミー変数行列を \mathbf{W} とし、選手位置行列を \mathbf{Z} とし、選手 i の質量の対数を y_i 、選手 i に対する説明変数行列を \mathbf{X}_i とすれば、

$$\begin{aligned} y_i &= \mathbf{W}\boldsymbol{\theta}_i + \mathbf{X}_i\boldsymbol{\beta}_i + \varepsilon_i (i = 1, \dots, n) \\ \mathbf{w}_t &\sim \text{Multi}(1, \boldsymbol{\pi}_t) (t = 1, \dots, T); \\ \pi_{kt} &= \frac{\exp\{z'_i\gamma_k\}}{\sum_{\ell=1}^K \exp\{z'_i\gamma_\ell\}} (k = 1, \dots, K) \end{aligned}$$

と、モデルを構築することが考えられる。ここで、 γ_ℓ, β_i はそれぞれボール位置を目的変数としたときの説明変数に対する回帰係数ベクトル、選手質量の対数を目的変数としたときの回帰係数ベクトルであり、 $\boldsymbol{\theta}_i$ は選手質量の対数を目的変数としたときのボール位置に対する回帰係数であり、 ε はそれぞれ誤差の分布である。また、 $\text{Multi}(1, \boldsymbol{\pi})$ は試行回数1、確率ベクトル $\boldsymbol{\pi}$ を持つ多項分布を表す。このようにモデル化することにより、選手位置によってボール位置が決定され、そのボール位置が選手質量に影響を与えているというモデルを構築することができる。これらのモデルと本稿で用いたモデルとの比較を当てはまりや予測の意味で比較する必要

がある。これらは今後の課題である。

また、本稿では、敵と味方との非対称性について、考察が十分にできていない。この非対称性についてデータからの意味付けを行うことが今後の課題である。

謝 辞

本研究で用いたデータはデータスタジアム株式会社より提供していただいたものである。ここに感謝申し上げます。また、査読者の方々より貴重なコメントとアドバイスをいただきました。心より感謝申し上げます。

参 考 文 献

- Dobson, S. and Goddard, J. (2008). Forecasting scores and result and testing the efficiency of the fixed-odds betting market in Scottish League Football, *Statistical Thinking in Sports* (eds. J. Albert and R. H. Koning), 91–119, CRC Press, New York, U.S.A.
- Ekin, A., Tekalp, A. M. and Mehrotra, R. (2003). Automatic soccer video analysis and summarization, *IEEE Transactions on Image Processing*, **12**, 796–807.
- 船山貴光, 山田実俊, 山本義郎, 内山秀一 (2016). 選手の動きによるチーム戦略と試合の流れの可視化, 「スポーツデータ解析における理論と事例に関する研究集会」, 統計数理研究所共同研究レポート, No. 363, 91–94.
- Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences, *Statistical Science*, **7**, 457–511.
- 神谷哲太, 中西航, 泉裕一郎 (2016). トラッキングデータを用いたサッカーの試合における戦況変化の抽出, 「スポーツデータ解析における理論と事例に関する研究集会」, 統計数理研究所共同研究レポート, No. 363, 77–82.
- 加藤久 (1999). サッカーの戦術とコンピュータ分析の現状と課題, オペレーションズ・リサーチ: 経営の科学, **44**, 125–131.
- Moura, F. A., Santana, E. J., Vieira, N. A., Santiago, P. R. P. and Cunha, A. S. (2015). Analysis of soccer players' positional variability during the 2012 UEFA European Championship: A case study, *Journal of Human Kinetics*, **47**, 225–236.
- Nadarajah, S. and Sundaraj, K. (2013). A survey on team strategies in robot soccer: Team strategies and role description, *Artificial Intelligence Review*, **40**, 271–304.
- 成塚拓真, 卯田純平, 山崎義弘 (2016). サッカーの対戦的特徴に現れる普遍的な統計性の探求, 「スポーツデータ解析における理論と事例に関する研究集会」, 統計数理研究所共同研究レポート, No. 363, 83–90.
- Ren, R. and Jose, J. M. (2005). Football video segmentation based on video production strategy, *Lecture Notes in Computer Science*, **3408**, 433–446.
- Rossi, P. E., Allenby, M. G. and McCulloch, R. (2005). *Bayesian Statistics and Marketing*, Wiley, Chichester, England.
- Rue, H. and Salvesen, O. (2000). Prediction and retrospective analysis of soccer matches in league, *Journal of the Royal Statistical Society: Series D*, **49**, 399–418.
- 瀧剛志, 長谷川純一 (1998). チームスポーツにおける集団行動解析のための特徴量とその応用, 電子情報通信学会論文誌, **81**, 1802–1811.
- Tinbergen, J. (1962). *Shaping the World Economy: Suggestions for an International Economic Policy*, Twentieth Century Fund, New York, U.S.A.

Quantitative Evaluation of Soccer Players' Movements

Jun Tsuchida¹ and Hiroshi Yadohisa²

¹Graduate School of Culture and Information Science, Doshisha University

²Department of Culture and Information Science, Doshisha University

Many studies have examined the movement of soccer players with the ball. Some studies have even investigated the movement of soccer players without the ball, but they tend to focus on evaluating the overall movements when executing gameplay strategies. Only a few studies have evaluated the soccer players themselves.

In this paper, we use player mass as an indicator to evaluate soccer players based on overall movements by a gravity model. In the gravity model, player mass is a parameter reflecting the movements of all players. Because the estimated player mass is equivalent to the parameters of a log-linear model, it corresponds to the main effect of the log-linear model. We calculated the density and distance among players to estimate the player mass for actual tracking data. Applying the estimated player mass and data from recorded gameplays to a Bayesian hierarchical model reveals the relationships between player mass and player movement.